

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 Date of Application: 2002年12月12日

出願番号 Application Number: 特願2002-360352

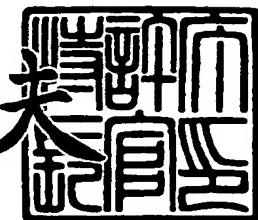
[ST. 10/C]: [JP2002-360352]

出願人 Applicant(s): 本田技研工業株式会社  
高根 定信

2003年11月13日

特許庁長官  
Commissioner,  
Japan Patent Office

今井 康夫



【書類名】 特許願

【整理番号】 H102355501

【提出日】 平成14年12月12日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/30

【発明者】

【住所又は居所】 東京都港区南青山2丁目1番1号

本田技研工業株式会社内

【氏名】 鈴木 雅晴

【発明者】

【住所又は居所】 東京都目黒区青葉台3-7-17 ソプラタワー190  
4号

株式会社組織活性化研究所内

【氏名】 高根 定信

【特許出願人】

【識別番号】 000005326

【氏名又は名称】 本田技研工業株式会社

【特許出願人】

【識別番号】 500545779

【氏名又は名称】 高根 定信

【代理人】

【識別番号】 100097113

【弁理士】

【氏名又は名称】 堀 城之

【手数料の表示】

【予納台帳番号】 044587

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0002986

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 情報処理装置および情報処理方法、並びに情報処理プログラム

【特許請求の範囲】

【請求項1】 テキストデータを入力する入力手段と、

前記テキストデータを記憶するテキストデータ記憶手段と、

前記テキストデータに対してワードカッティング処理を実行するワードカッティング手段と、

前記ワードカッティング処理が行われた前記テキストデータに対して構文解析処理を行う構文解析手段と、

前記構文解析処理が行われた前記テキストデータからシソーラスを作成するシソーラス作成手段と、

前記シソーラス作成手段によって作成された前記シソーラスを記憶するシソーラス記憶手段と、

前記ワードカッティングおよび前記構文解析が行われた前記テキストデータに対して仕分け処理を行うシソーラス仕分け手段と、

前記シソーラス仕分け手段による仕分け結果を記憶する仕分け結果記憶手段と、

前記仕分け結果記憶手段によって記憶されている前記仕分け結果に基づいて、各シソーラス毎に出現頻度を算出する出現頻度算出手段と、

前記出現頻度算出手段によって算出された結果を記憶する出現頻度記憶手段と、

シソーラス間の相関係数を算出する相関係数算出手段と、

相関係数算出手段によって算出されたシソーラス間の相関係数を記憶する相関係数記憶手段と、

シソーラス毎の相関係数の合計を算出するシソーラス毎の相関係数合計算出手段と、

シソーラス毎の相関係数合計算出手段によって算出されたシソーラス毎の相関係数の合計を記憶するシソーラス毎の相関係数合計記憶手段と、

出現頻度記憶手段によって記憶されている出現頻度と、シソーラス毎の相関係

数合計記憶手段によって記憶されているシソーラス毎の相関係数合計に基づいて、グラフを作成し、表示するグラフ作成表示手段と  
を備え、

前記ワードカッティング手段および前記構文解析手段は、前記シソーラス作成手段によって作成された前記シソーラスに基づいて再度ワードカッティング処理および構文解析処理を行う

ことを特徴とする情報処理装置。

【請求項2】 テキストデータを入力する入力ステップと、  
前記テキストデータを記憶するテキストデータ記憶ステップと、  
前記テキストデータに対してワードカッティング処理を実行するワードカッティングステップと、

前記ワードカッティング処理が行われた前記テキストデータに対して構文解析処理を行う構文解析ステップと、

前記構文解析処理が行われた前記テキストデータからシソーラスを作成するシソーラス作成ステップと、

前記シソーラス作成ステップにおいて作成された前記シソーラスを記憶するシソーラス記憶ステップと、

前記シソーラス記憶ステップにおいて記憶された前記シソーラスに基づいて、再度ワードカッティング処理および構文解析処理を行うワードカッティングおよび構文解析ステップと、

前記ワードカッティングおよび前記構文解析が行われた前記テキストデータに対して仕分け処理を行うシソーラス仕分けステップと、

前記シソーラス仕分けステップにおける仕分け結果を記憶する仕分け結果記憶ステップと、

前記仕分け結果記憶ステップにおいて記憶された前記仕分け結果に基づいて、各シソーラス毎に出現頻度を算出する出現頻度算出ステップと、

前記出現頻度算出ステップにおいて算出された結果を記憶する出現頻度記憶ステップと、

シソーラス間の相関係数を算出する相関係数算出ステップと、

相関係数算出ステップにおいて算出されたシソーラス間の相関係数を記憶する  
相関係数記憶ステップと、

シソーラス毎の相関係数の合計を算出するシソーラス毎の相関係数合計算出ス  
テップと、

シソーラス毎の相関係数合計算出ステップにおいて算出されたシソーラス毎の  
相関係数の合計を記憶するシソーラス毎の相関係数合計記憶ステップと、

出現頻度記憶ステップにおいて記憶された出現頻度と、シソーラス毎の相関係  
数合計記憶ステップにおいて記憶されたシソーラス毎の相関係数合計とに基づい  
て、グラフを作成し、表示するグラフ作成表示ステップと  
を備えることを特徴とする情報処理方法。

【請求項3】 テキストデータを入力する入力ステップと、  
前記テキストデータを記憶するテキストデータ記憶ステップと、  
前記テキストデータに対してワードカッティング処理を実行するワードカッテ  
ィングステップと、  
前記ワードカッティング処理が行われた前記テキストデータに対して構文解析  
処理を行う構文解析ステップと、  
前記構文解析処理が行われた前記テキストデータからシソーラスを作成するシ  
ソーラス作成ステップと、  
前記シソーラス作成ステップにおいて作成された前記シソーラスを記憶するシ  
ソーラス記憶ステップと、  
前記シソーラス記憶ステップにおいて記憶された前記シソーラスに基づいて、  
再度ワードカッティング処理および構文解析処理を行うワードカッティングおよ  
び構文解析ステップと、  
前記ワードカッティングおよび前記構文解析が行われた前記テキストデータに  
対して仕分け処理を行うシソーラス仕分けステップと、  
前記シソーラス仕分けステップにおける仕分け結果を記憶する仕分け結果記憶  
ステップと、  
前記仕分け結果記憶ステップにおいて記憶された前記仕分け結果に基づいて、  
各シソーラス毎に出現頻度を算出する出現頻度算出ステップと、

前記出現頻度算出ステップにおいて算出された結果を記憶する出現頻度記憶ステップと、

シソーラス間の相関係数を算出する相関係数算出ステップと、

相関係数算出ステップにおいて算出されたシソーラス間の相関係数を記憶する相関係数記憶ステップと、

シソーラス毎の相関係数の合計を算出するシソーラス毎の相関係数合計算出ステップと、

シソーラス毎の相関係数合計算出ステップにおいて算出されたシソーラス毎の相関係数の合計を記憶するシソーラス毎の相関係数合計記憶ステップと、

出現頻度記憶ステップにおいて記憶された出現頻度と、シソーラス毎の相関係数合計記憶ステップにおいて記憶されたシソーラス毎の相関係数合計とに基づいて、グラフを作成し、表示するグラフ作成表示ステップと  
をコンピュータに実行させることを特徴とする情報処理プログラム。

#### 【発明の詳細な説明】

##### 【0001】

##### 【発明の属する技術分野】

本発明は、情報処理装置および情報処理方法、並びに情報処理プログラムに関し、特に、テキストデータを元にシソーラスを作成し、シソーラス間の相関を求める情報処理装置および情報処理方法、並びに情報処理プログラムに関する。

##### 【0002】

##### 【従来の技術】

各単語の出現回数の集計結果を参照して、切り出した単語から出現頻度が規定値以上の単語を抽出し、その抽出した各単語間の関連性を評価して、関連度が規定値以上の共起単語のクラスタを生成する。このとき、予め、分析対象のテキストに合わせてカテゴリ辞書を作成すれば、そのテキストの分析結果を提示することができるというものがある（例えば、特許文献1参照）。

##### 【0003】

##### 【特許文献1】

特開2001-101194号公報（図1）

**【0004】****【発明が解決しようとする課題】**

しかしながら、テキストデータから抽出したキーワード同士の相関に基づいて、テキストデータの特徴を検出することができないという問題があった。

**【0005】**

本発明はこのような状況に鑑みてなされたものであり、テキストデータから抽出したキーワード同士の相関に基づいて、テキストデータの特徴を検出することができるようとするものである。

**【0006】****【課題を解決するための手段】**

請求項1に記載の情報処理装置は、テキストデータを入力する入力手段と、テキストデータを記憶するテキストデータ記憶手段と、テキストデータに対してワードカッティング処理を実行するワードカッティング手段と、ワードカッティング処理が行われたテキストデータに対して構文解析処理を行う構文解析手段と、構文解析処理が行われたテキストデータからシソーラスを作成するシソーラス作成手段と、シソーラス作成手段によって作成されたシソーラスを記憶するシソーラス記憶手段と、ワードカッティングおよび構文解析が行われたテキストデータに対して仕分け処理を行うシソーラス仕分け手段と、シソーラス仕分け手段による仕分け結果を記憶する仕分け結果記憶手段と、仕分け結果記憶手段によって記憶されている仕分け結果に基づいて、各シソーラス毎に出現頻度を算出する出現頻度算出手段と、出現頻度算出手段によって算出された結果を記憶する出現頻度記憶手段と、シソーラス間の相関係数を算出する相関係数算出手段と、相関係数算出手段によって算出されたシソーラス間の相関係数を記憶する相関係数記憶手段と、シソーラス毎の相関係数の合計を算出するシソーラス毎の相関係数合計算出手段と、シソーラス毎の相関係数合計算出手段によって算出されたシソーラス毎の相関係数の合計を記憶するシソーラス毎の相関係数合計記憶手段と、出現頻度記憶手段によって記憶されている出現頻度と、シソーラス毎の相関係数合計算出手段によって記憶されているシソーラス毎の相関係数合計に基づいて、グラフを作成し、表示するグラフ作成表示手段とを備え、ワードカッティング手段お

および構文解析手段は、シソーラス作成手段によって作成されたシソーラスに基づいて再度ワードカッティング処理および構文解析処理を行うことを特徴とする。

請求項2に記載の情報処理方法は、テキストデータを入力する入力ステップと、テキストデータを記憶するテキストデータ記憶ステップと、テキストデータに対してワードカッティング処理を実行するワードカッティングステップと、ワードカッティング処理が行われたテキストデータに対して構文解析処理を行う構文解析ステップと、構文解析処理が行われたテキストデータからシソーラスを作成するシソーラス作成ステップと、シソーラス作成ステップにおいて作成されたシソーラスを記憶するシソーラス記憶ステップと、シソーラス記憶ステップにおいて記憶されたシソーラスに基づいて、再度ワードカッティング処理および構文解析処理を行なうワードカッティングおよび構文解析ステップと、ワードカッティングおよび構文解析が行われたテキストデータに対して仕分け処理を行うシソーラス仕分けステップと、シソーラス仕分けステップにおける仕分け結果を記憶する仕分け結果記憶ステップと、仕分け結果記憶ステップにおいて記憶された仕分け結果に基づいて、各シソーラス毎に出現頻度を算出する出現頻度算出ステップと、出現頻度算出ステップにおいて算出された結果を記憶する出現頻度記憶ステップと、シソーラス間の相関係数を算出する相関係数算出ステップと、相関係数算出ステップにおいて算出されたシソーラス間の相関係数を記憶する相関係数記憶ステップと、シソーラス毎の相関係数の合計を算出するシソーラス毎の相関係数合計算出ステップと、シソーラス毎の相関係数合計算出ステップにおいて算出されたシソーラス毎の相関係数の合計を記憶するシソーラス毎の相関係数合計記憶ステップと、出現頻度記憶ステップにおいて記憶された出現頻度と、シソーラス毎の相関係数合計記憶ステップにおいて記憶されたシソーラス毎の相関係数合計とに基づいて、グラフを作成し、表示するグラフ作成表示ステップとを備えることを特徴とする。

請求項3に記載の情報処理プログラムは、テキストデータを入力する入力ステップと、テキストデータを記憶するテキストデータ記憶ステップと、テキストデータに対してワードカッティング処理を実行するワードカッティングステップと、ワードカッティング処理が行われたテキストデータに対して構文解析処理を行

う構文解析ステップと、構文解析処理が行われたテキストデータからシソーラスを作成するシソーラス作成ステップと、シソーラス作成ステップにおいて作成されたシソーラスを記憶するシソーラス記憶ステップと、シソーラス記憶ステップにおいて記憶されたシソーラスに基づいて、再度ワードカッティング処理および構文解析処理を行うワードカッティングおよび構文解析ステップと、ワードカッティングおよび構文解析が行われたテキストデータに対して仕分け処理を行うシソーラス仕分けステップと、シソーラス仕分けステップにおける仕分け結果を記憶する仕分け結果記憶ステップと、仕分け結果記憶ステップにおいて記憶された仕分け結果に基づいて、各シソーラス毎に出現頻度を算出する出現頻度算出ステップと、出現頻度算出ステップにおいて算出された結果を記憶する出現頻度記憶ステップと、シソーラス間の相関係数を算出する相関係数算出ステップと、相関係数算出ステップにおいて算出されたシソーラス間の相関係数を記憶する相関係数記憶ステップと、シソーラス毎の相関係数の合計を算出するシソーラス毎の相関係数合計算出ステップと、シソーラス毎の相関係数合計算出ステップにおいて算出されたシソーラス毎の相関係数の合計を記憶するシソーラス毎の相関係数合計記憶ステップと、出現頻度記憶ステップにおいて記憶された出現頻度と、シソーラス毎の相関係数合計記憶ステップにおいて記憶されたシソーラス毎の相関係数合計とにに基づいて、グラフを作成し、表示するグラフ作成表示ステップとをコンピュータに実行させることを特徴とする。

### 【0007】

#### 【発明の効果】

以上の如く、本発明に係る情報処理装置および情報処理方法、並びに情報処理プログラムによれば、テキストデータを入力する入力ステップと、テキストデータを記憶するテキストデータ記憶ステップと、テキストデータに対してワードカッティング処理を実行するワードカッティングステップと、ワードカッティング処理が行われたテキストデータに対して構文解析処理を行う構文解析ステップと、構文解析処理が行われたテキストデータからシソーラスを作成するシソーラス作成ステップと、シソーラス作成ステップにおいて作成されたシソーラスを記憶するシソーラス記憶ステップと、シソーラス記憶ステップにおいて記憶されたシ

ソーラスに基づいて、再度ワードカッティング処理および構文解析処理を行うワードカッティングおよび構文解析ステップと、ワードカッティングおよび構文解析が行われたテキストデータに対して仕分け処理を行うシソーラス仕分けステップと、シソーラス仕分けステップにおける仕分け結果を記憶する仕分け結果記憶ステップと、仕分け結果記憶ステップにおいて記憶された仕分け結果に基づいて、各シソーラス毎に出現頻度を算出する出現頻度算出ステップと、出現頻度算出ステップにおいて算出された結果を記憶する出現頻度記憶ステップと、シソーラス間の相関係数を算出する相関係数算出ステップと、相関係数算出ステップにおいて算出されたシソーラス間の相関係数を記憶する相関係数記憶ステップと、シソーラス毎の相関係数の合計を算出するシソーラス毎の相関係数合計算出ステップと、シソーラス毎の相関係数合計算出ステップにおいて算出されたシソーラス毎の相関係数の合計を記憶するシソーラス毎の相関係数合計記憶ステップと、出現頻度記憶ステップにおいて記憶された出現頻度と、シソーラス毎の相関係数合計記憶ステップにおいて記憶されたシソーラス毎の相関係数合計に基づいて、グラフを作成し、表示するようにしたので、テキストデータから抽出したキーワードから作成したシソーラス同士の相関および出現頻度に基づいて、テキストデータの特徴を検出し、テキストデータに隠れている潜在的な意味を類推することが可能となる。

### 【0008】

#### 【発明の実施の形態】

図1は、本発明の情報処理装置の一実施の形態の機能ブロック図である。本実施の形態は、パーソナルコンピュータ等で構成されている。同図に示すように、本実施の形態は、機能的に次のようなブロックで構成される。各ブロックの処理は、実際には所定のアプリケーションプログラムによって実行され、各記憶部は、図示せぬハードディスクによって実現される。

### 【0009】

各ブロックの機能を簡単に説明する。入力部1は、テキストデータを入力し、テキスト記憶部2に記憶させるようになっている。ワードカッティング部3は、テキスト記憶部2に記憶されているテキストデータに対して、ワードカッティン

グ処理を実行するようになっている。構文解析部4は、ワードカッティング処理されたテキストデータに対して構文解析を行うようになっている。

#### 【0010】

シソーラス作成部5は、テキスト記憶部2に記憶されているテキストデータからシソーラスを作成するようになっている。シソーラス記憶部6は、作成されたシソーラスを記憶するようになっている。シソーラス仕分け部7は、全サンプルについて、シソーラス毎に仕分け処理を行うようになっている。仕分け結果記憶部8は、仕分け結果を記憶するようになっている。出現頻度算出部9は、仕分け結果記憶部8に記憶されているデータに基づいて、各シソーラス毎に出現頻度を算出するようになっている。出現頻度記憶部10は、出現頻度算出部9によって算出された結果を記憶するようになっている。

#### 【0011】

相関係数算出部11は、シソーラス間の相関係数を算出するようになっている。相関係数記憶部12は、相関係数算出部11において算出された相関係数を記憶するようになっている。シソーラス毎の相関係数合計算出部13は、求めた相関係数をシソーラス毎に合計するようになっている。シソーラス毎の相関係数合計記憶部14は、シソーラス毎の相関係数合計算出部13において算出されたシソーラス毎の相関係数の合計を記憶するようになっている。グラフ作成表示部15は、出現頻度記憶部10に記憶されている出現頻度と、シソーラス毎の相関係数合計記憶部14に記憶されているシソーラス毎の相関係数合計とに基づいて、グラフを作成し、表示するようになっている。

#### 【0012】

次に、図2のフローチャートを参照して、本実施の形態の処理手順について説明する。ここでは、顧客から寄せられる苦情や要望等のテキストデータを分析する場合について説明する。まず最初に、ステップS1において、入力部1より顧客毎にテキストデータが入力される。例えば、ある顧客から「先週、部品を注文したのだが、まだ納入されない。」が入力されたものとする。入力されたテキストデータは、テキスト記憶部2に記憶される。

#### 【0013】

次に、ステップS2において、ワードカッティング部3により、所定のテキストマイニングツール（アプリケーションソフトウェア）によるワードカッティング処理が行われる。例えば、上記テキストは、「先週、部品を注文したのだが、まだ納入されない。」となる。

#### 【0014】

次に、ステップS3において、構文解析部4により、テキストマイニングツールによる構文解析処理が行われる。例えば、上記テキストは、「先週、部品を注文したのだが、まだ納入されない。」となる。

#### 【0015】

次に、ステップS4において、シソーラス作成部5により、類義語（キーワード）を集約したシソーラスの作成が行われる。例えば、図3に示すように、「先週」というシソーラスに、「一週間」等の類義語のキーワードを集約する。また、「注文する」というシソーラスに、「オーダーしたが」等のキーワードを集約する。また、「納入する」というシソーラスに、「搬入する」等のキーワードを集約する。また、「部品」というシソーラスに、「パーツ」等のキーワードを集約する。また、「連絡が」というシソーラスに、「情報を」等のキーワードを集約する。作成されたシソーラスは、シソーラス記憶部6に記憶される。

#### 【0016】

次に、ステップS5において、いま作成され、シソーラス記憶部6に記憶されたシソーラスに基づいて、ワードカッティング部3により、再度、ワードカッティング処理が行われ、構文解析部4により、再度、構文解析処理が行われる。

#### 【0017】

次に、ステップS6において、シソーラス仕分け部7により、全顧客からのテキストデータに対して、その内容について、各シソーラス毎に仕分けが行われる。例えば、各顧客毎に、その顧客の苦情等のテキストデータに含まれるシソーラスには「1」を、含まれないシソーラスには「0」をセットする。仕分け結果は仕分け結果記憶部8に記憶される。

#### 【0018】

図4は、仕分け結果記憶部8に記憶されている仕分け結果を示している。図中

、「K-1」、「K-2」、「K-3」、…、「K-n」は、顧客を識別する識別番号を示している。この例では、例えば、顧客K-1により、「注文する」、「部品」というシソーラスに含まれるキーワードを含むテキストデータが入力されたことがわかる。

#### 【0019】

次に、ステップS7において、相関係数算出部11により、シソーラス間の相関係数が求められる。例えば、「注文する」と、「納入する」の相関係数は、次式で表される。

#### 【0020】

$$\text{相関係数 } r_{\text{注文}-\text{納入}} = (S_{\text{注文}-\text{納入}}) / (S_{\text{注文}} \cdot S_{\text{納入}})$$

ただし、 $S_{\text{注文}-\text{納入}}$  は、共分散、 $S_{\text{注文}}$ 、 $S_{\text{納入}}$  は、それぞれ標準偏差である。

#### 【0021】

$$\text{共分散 } S_{\text{注文}-\text{納入}} = ((\text{注文}_1 - \bar{\text{注文}})(\text{納入}_1 - \bar{\text{納入}}) + (\text{注文}_2 - \bar{\text{注文}})(\text{納入}_2 - \bar{\text{納入}}) + \dots + (\text{注文}_n - \bar{\text{注文}})(\text{納入}_n - \bar{\text{納入}})) / (n - 1)$$

#### 【0022】

$$\begin{aligned} \text{標準偏差 } S_{\text{注文}} &= ((\text{注文}_1 - \bar{\text{注文}})^2 + (\text{注文}_2 - \bar{\text{注文}})^2 + \dots \\ &+ (\text{注文}_n - \bar{\text{注文}})^2) / (n - 1) \end{aligned}$$

$$\begin{aligned} \text{標準偏差 } S_{\text{納入}} &= ((\text{納入}_1 - \bar{\text{納入}})^2 + (\text{納入}_2 - \bar{\text{納入}})^2 + \dots \\ &+ (\text{納入}_n - \bar{\text{納入}})^2) / (n - 1) \end{aligned}$$

#### 【0023】

以下、同様にして、全シソーラス間の相関係数を求め、相関係数記憶部12に記憶させる。図5は、シソーラス間の相関係数を示している。例えば、シソーラス「先週」とシソーラス「注文する」の間の相関係数は0.025となっている。同一のシソーラス間の相関係数は1となる。

#### 【0024】

次に、ステップS8において、シソーラス毎の相関係数合計算出部13により、ステップS7で求められ、相関係数記憶部12に記憶された相関係数を、各シ

ソーラス毎に合計する。例えば、シソーラス「先週」の場合、 $1 + 0.025 + 0.038 + 0.001 + \dots$ となる。このとき、同一のシソーラス間の相関係数1を除く。以下同様にして、シソーラス「注文する」、「納入する」、「部品」等の他のシソーラスについても、相関係数の合計を求める。求められた各シソーラス毎の相関係数の合計は、シソーラス毎の相関係数合計記憶部14に記憶される。

### 【0025】

次に、ステップS9において、出現頻度算出部9により、各シソーラスの出現頻度が求められる。即ち、図6に示すように、シソーラス毎の仕分け結果（図4）に基づいて、各シソーラス毎の出現頻度が求められる。図6の例では、例えば、シソーラス「先週」については、顧客K-2、K-3、…K-nの苦情等のテキストデータの中に、そのシソーラス「先週」が含まれていることがわかる。シソーラス「先週」についての各顧客毎の値を合計することにより、出現数Aが算出される。同様に、シソーラス「注文」の出現数はB、シソーラス「納入する」の出現数はC、シソーラス「部品」の出現数は「D」となる。全シソーラスの出現数の合計 $\Sigma (A + B + C + D + \dots)$ を求め、各シソーラスの出現頻度をパーセンテージで表す。

### 【0026】

例えば、シソーラス「先週」の出現頻度は、 $(A / \Sigma (A + B + C + D + \dots)) * 100$  (%)となる。算出された各シソーラス毎の出現頻度は、出現頻度記憶部10に記憶される。

### 【0027】

次に、ステップS10において、グラフ作成表示部15により、各シソーラスの出現頻度（%）をx軸、シソーラス毎の相関係数の合計をy軸にとり、プロットして、グラフが作成される。図7は、作成されたグラフを示している。

### 【0028】

図7に示すように、各顧客が苦情等で発言した中で、出現回数は多くはないが、他のシソーラスとの結びつき（何らかの意味上の関連性）が強いシソーラスが第1のグループに現れる。

**【0029】**

また、他のシソーラスとの結びつきはそれほど強くないが、出現回数が多いもの。即ち、頻繁に発言されるので見逃せないシソーラスが第3のグループに現れる。

**【0030】**

ここでは、相関係数の数値の大小に意味を求めるのではなく、ある一定レベルに基準を定め、その基準レベルを超えたものは結びつきが強く、下回ったものは結びつきが弱いものと判断する。

**【0031】**

y座標の数値がある一定レベルを超えたか否かが重要であり、一定のレベルを超えたシソーラスは、他のキーワードと結びついて何らかの意味を構成している可能性が高いとした。このケースの場合、テキストは顧客の部品に関する苦情であるので、「何らかの意味」とは苦情、即ち「潜在的な不満」を示すものであると類推できる。

**【0032】**

なお、上記実施の形態の構成及び動作は例であって、本発明の趣旨を逸脱しない範囲で適宜変更することができることは言うまでもない。

**【0033】****【図面の簡単な説明】****【図1】**

本発明の情報処理装置の一実施の形態の機能ブロック図である。

**【図2】**

本実施の形態の処理手順を説明するためのフローチャートである。

**【図3】**

類義語を集約したシソーラスの例を示す図である。

**【図4】**

シソーラス毎の仕分け結果を示す図である。

**【図5】**

シソーラス毎の相関係数を示す図である。

**【図6】**

各シソーラス毎の出現頻度を示す図である。

**【図7】**

各シソーラスの出現頻度と相関係数との関係を示すグラフである。

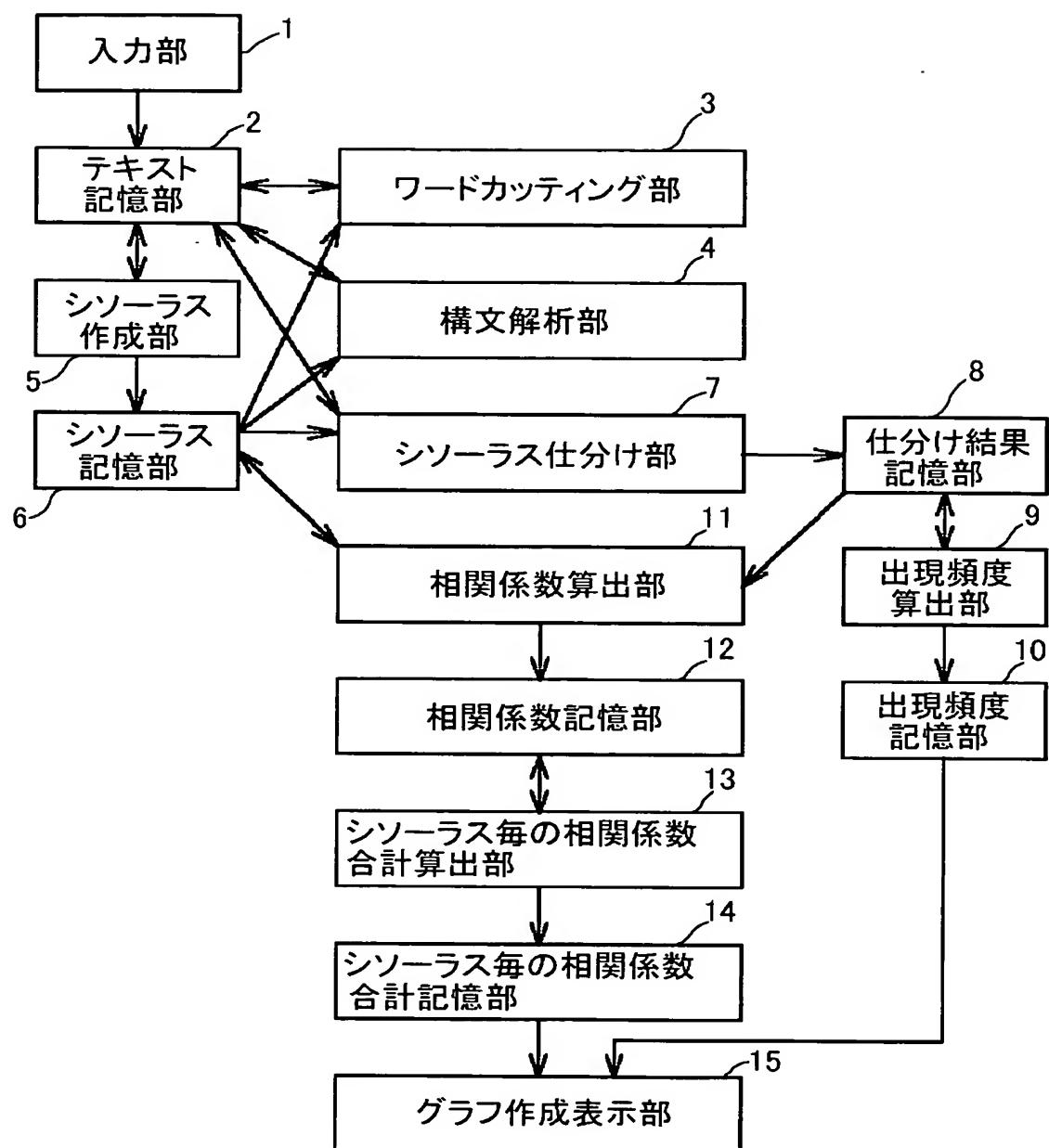
**【符号の説明】**

- 1 入力部
- 2 テキスト記憶部
- 3 ワードカッティング部
- 4 構文解析部
- 5 シソーラス作成部
- 6 シソーラス記憶部
- 7 シソーラス仕分け部
- 8 仕分け結果記憶部
- 9 出現頻度算出部
- 10 出現頻度記憶部
- 11 相関係数算出部
- 12 相関係数記憶部
- 13 シソーラス毎の相関係数合計算出部
- 14 シソーラス毎の相関係数合計記憶部
- 15 グラフ作成表示部

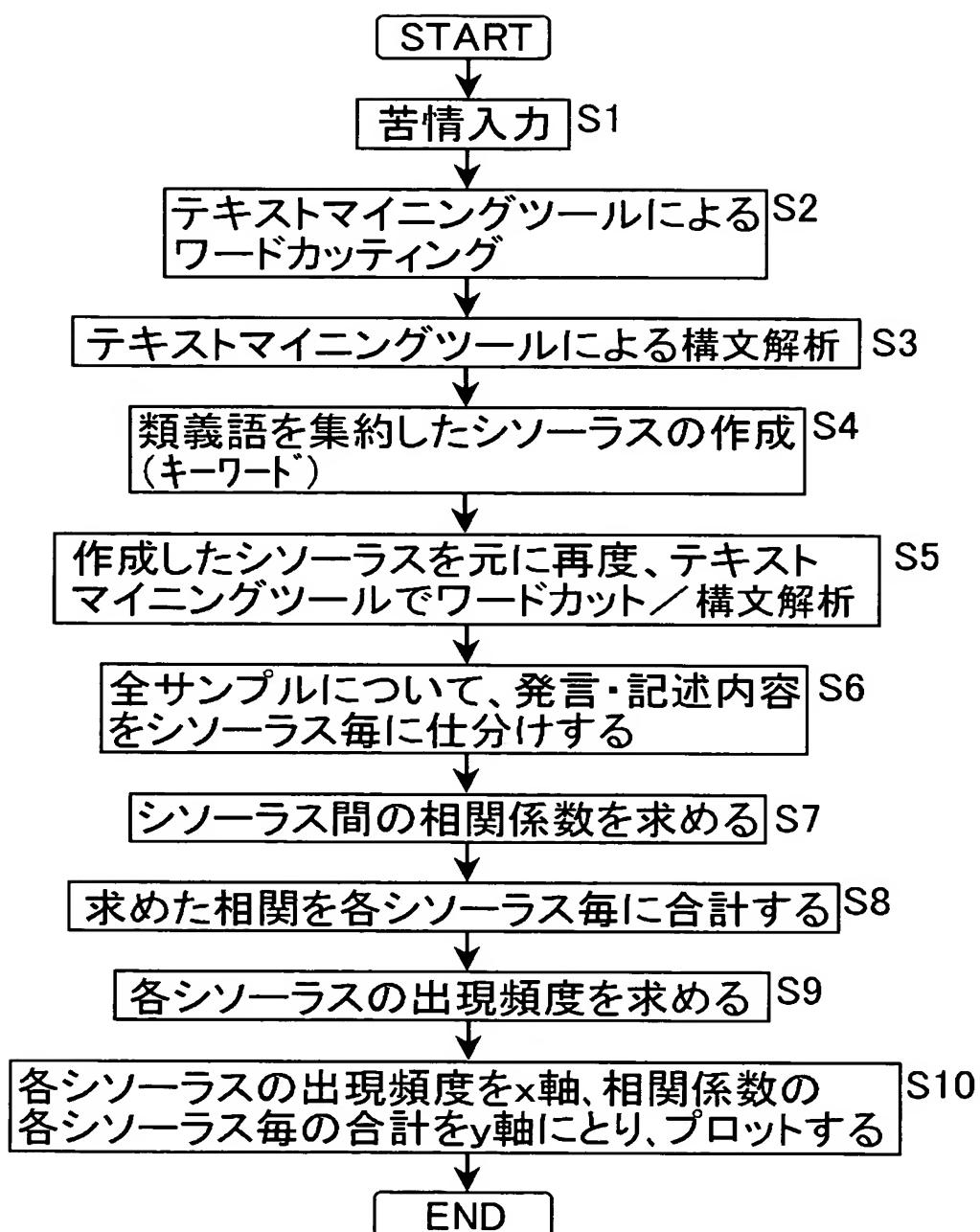
【書類名】

図面

【図 1】



【図2】



【図 3】

先週	注文する	納入する	部品	連絡が	---
一週前	オーダーしたが	搬入する	パーツ	情報を	---
---	---	---	---	---	---

【図 4】

	先週	注文する	納入する	部品	-----
K-1	0	1	0	1	-----
K-2	1	1	1	0	-----
K-3	1	1	0	0	-----
---	---	---	---	---	---
K-n	1	1	0	1	-----

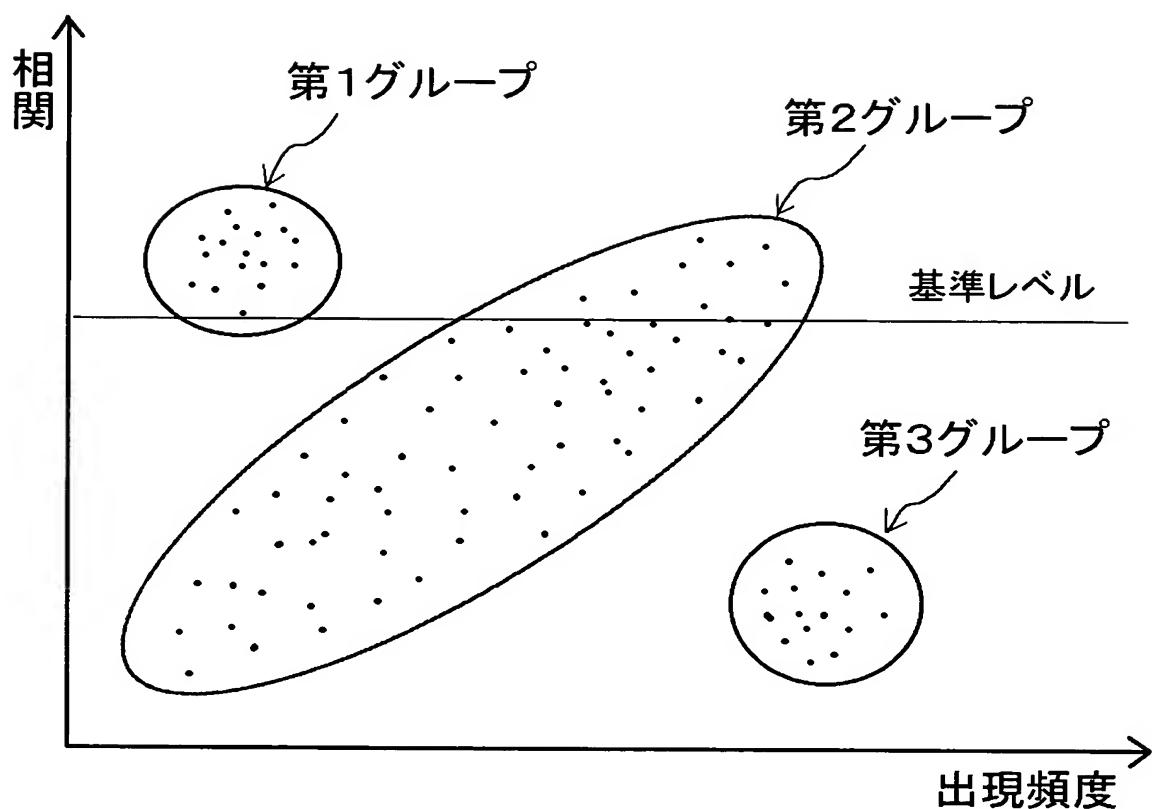
【図 5】

	先週	注文する	納入する	部品	-----
先週	1	0.025	0.038	0.001	-----
注文する	0.025	1	0.064	0.021	-----
納入する	0.038	0.064	1	0.004	-----
部品	0.001	0.021	0.004	1	-----
⋮	⋮	⋮	⋮	⋮	⋮
合計					

【図 6】

	先週	注文する	納入する	部品	-----
K-1	0	1	0	1	-----
K-2	1	1	1	0	-----
K-3	1	1	0	0	-----
⋮	⋮	⋮	⋮	⋮	⋮
K-n	1	1	0	1	-----
合計	A	B	C	D	-----
					$\Sigma (A+B+C+D\dots)$

【図7】



【書類名】 要約書

【要約】

【課題】 テキストデータの特徴を検出し、テキストデータに隠れている潜在的な意味を類推することができるようとする。

【解決手段】 入力部1から入力されたテキストデータに対して、ワードカッティング部3がワードカッティング処理し、構文解析部4が構文解析し、その結果からシソーラス作成部5がシソーラスを作成し、再ワードカッティング・構文解析後、シソーラス仕分け部7が仕分けし、出現頻度算出部がシソーラスの出現頻度を算出し、相関係数算出部11がシソーラス間の相関係数を算出し、シソーラス毎の相関係数合計算出部13がシソーラス毎の相関係数の合計を算出し、グラフ作成表示部15が出現頻度とシソーラス毎の相関係数の合計に基づいてグラフを作成する。

【選択図】 図1

特願2002-360352

出願人履歴情報

識別番号 [000005326]

1. 変更年月日 1990年 9月 6日

[変更理由] 新規登録

住所 東京都港区南青山二丁目1番1号  
氏名 本田技研工業株式会社

特願 2002-360352

出願人履歴情報

識別番号 [500545779]

1. 変更年月日 2001年12月 4日  
[変更理由] 住所変更  
住 所 東京都目黒区青葉台3-7-17 ソプラタワー1904号  
氏 名 高根 定信